Bayesianism, the Methodology of Scientific Research Programmes, and Duhem's Problem.

Jon Dorling,
History and Philosophy of Science,
University of London
Chelsea College,
Manresa Road, London SW3, U.K.

Abstract

The detailed analysis of a particular numerical example is used to illustrate the way in which a Bayesian approach to scientific inference solves the Duhemian problem of which of a conjunction of hypotheses to reject when they jointly yield a prediction which is refuted. The same example shows, in agreement with the views of Lakatos, that a refutation need have little effect on a scientist's confidence in the hard core of a successful research programme, when a comparable confirmation would greatly enhance that confidence. The numerical results are very striking and comparatively stable against variations in the initial probabilities and likelihoods, which are set up so as to do approximate justice to an actual historical case.

Bayesianism, the Methodology of Research Programmes, and Duhem's Problem.

Jon Dorling,
Department of Hist. and Phil. of Sci.,
Chelsea College, Univ. of London,
Manresa Road, London SW3, U.K.

The derivation of any particular scientific prediction will depend on a large number of hypotheses and auxiliary hypotheses. If the prediction is falsified a problem will then arise as to which of these numerous hypotheses we should reject. The statement of this problem goes back at least as far as Duhem and I shall call it Duhem's problem. I shall show how a Bayesian probabilistic approach to scientific inference solves this problem.

In Imre Lakatos's Methodolggy of Scientific Research Programmes a typical case of this problem arises when a prediction depends jointly on some hypothesis constituting part of the 'hard core' of a successful research programme and on one or more auxiliary hypotheses constituting part of the 'protective belt'. Lakatos's view is that if such a prediction is refuted then a scientist should reject the hypothesis belonging to the protective belt but hang on to the hypothesis belonging to the hard core, and indeed that his faith in the hard core of the research programme need be little affected. This approach leads to a striking asymmetry between the effects of verification and falsification of predictions, verifications counting for a great deal, on Lakatos's view, and falsifications counting for very little. This asymmetry, which is in the opposite direction to what the Popperian tradition might have led us to expect, does, nevertheless, seem to square well with the actual behaviour of scientists in many instances. I shall show that just this asymmetry is, surprisingly enough, to be expected from a Bayesian analysis of the problem.

Let T be the hard core of the theory, H the auxiliary hypothesis, and E the experimental prediction. To take a specific example, let T be the hard core of mid-nineteenth century coelestial mechanics, let H be the auxiliary hypothesis that the effects of tidal friction are not of a sufficient order of magnitude to appreciably affect the phenomena that we are interested in, and let E be a prediction of the amount of the moon's secular acceleration.

Suppose, at the time in question, T has an initial subjective probability of 0.9, and H an initial subjective probability of 0.6:

$$p(T) = .9 \qquad p(H) = .6.$$

Suppose also that E is entailed by T and H; then

$$p(E,T.H) = 1.$$

Now let E' be the astronomically observed secular acceleration of the moon. If E' is not equal to E, then

$$p(E',T.H) = 0$$

I am interested in computing p(T,E') and p(H,E'). By Bayes's theorem

$$p(T,E') = p(E',T)p(T)/p(E')$$
$$p(H,E') = p(E',H)p(H)/p(E').$$

I assume T is irrelevant to H: p(H.T)=p(H)p(T); hence I can assert

$$p(E',T) = p(E',T.H)p(H) + p(E',T.-H)p(-H)$$
$$p(E',H) = p(E',T.H)p(T) + p(E',-T.H)p(-T)$$
$$p(E') = p(E',T)p(T) + p(E',-T)p(-T).$$

ace for new paragraph

To solve my problem I need now additional subjective information about the values of p(E',T.-H), p(E',-T.H), and p(E',-T) = p(E',-T.H)p(H) + p(E',-T.-H)p(-H). Now in a first draft of this paper I conjectured that the values of all three of these conditional subjective probabilities could be taken as very small, say approximately 0.001, on the grounds that the value of E' was a surprising astronomical discovery, unexpected on any theory when it was first made (in the mid-eighteenth century) and known to about two significant figures by the mid-nineteenth century. I still think this is roughly right - for the purposes of my general philosophical argument I just want E' to be some experimental fact, which is a priori unlikely in the sense that no plausible rival theory, to the one that we hope will explain it, predicts E' either quantitatively or even qualitatively – but I now think that it would be an over-simplification to take the three conditional probabilities in question as equal. $p(E',T.-H)$ depends on the odds at which one would wager on the particular observational result E', if no astronomical determination of the existence or magnitude of such an effect had yet been made, and if one believed in the hard core of Newtonian coelestial mechanics, and also believed that tidal friction might well produce an effect of the order of magnitude of E'. It seems that odds of about 1/20 might be about right here, so I shall set

$$p(E',T.-H) = 0.05.$$

Space for new paragraph

$p(E',-T.H)$ corresponds, on the other hand, to the case where I assume that I am convinced, for some independent reason, that the hard core of the Newtonian theory is false, but that I am also convinced that tidal friction could not produce an effect of the order of magnitude of the purely hypothetical value E', concerning which I am invited to wager.

$$p(E',-T.H) = 0.001$$

seems a fair estimate for this case.

Space for new paragraph

$p(E',-T)$ is now determined provided we can estimate $p(E',-T.-H)$. Presumably the latter must also be taken as 1/20 since the previous argument for that figure did not really depend on the belief in T but only on the belief in -H. Hence we obtain

$$\begin{aligned} p(E',-T) &= p(E',-T.H)p(H) + p(E',-T.-H)p(-H) \\ &= 0.001 \text{ x } .6 \quad + \quad 0.05 \text{ x } .4 \\ &= 0.0206. \end{aligned}$$

Space for new paragraph

Substituting these figures in the previous formulae, we find

$$p(E',T) = 0.02, \quad p(E',H) = 0.0001, \quad p(E') = 0.02006,$$

and hence, by Bayes's theorem,

$$p(T,E') = 0.02 \text{ x } .9/0.02006 = .8976$$

$$p(H,E') = 0.0001 \text{ x } .6/0.02006 = .003.$$

Space for new paragraph

This result shows not merely that a thoroughgoingly Bayesian approach does indeed solve the Duhem problem of how to apportion the blame for a predictive failure, but given the not implausible input of subjective initial probabilities and likelihoods which I have chosen for this example, the results are astonishingly asymmetric and in line with the dogmatic recommendations of Lakatos to a quite unexpected degree. Nor is this a consequence of sleight-of-hand in my choice of initial probabilities and likelihoods. Provided only that E' is no more readily explainable by any plausible rival theory to T, and provided T starts off more probable than not and substantially more probable than H, then whatever actual numbers one inserts, one obtains a qualitatively similar result to the one I have just derived.

It is interesting to look as well at the alternative case where E' turns out to be equal to E, that is to say where the prediction of the theory is confirmed rather than refuted. Here $p(E',T.H) = 1$, and computation, assuming the same initial probabilities and likelihoods ~~otherwise~~,yields:

$$p(E',T) = 0.62, \quad p(E',H) = 0.9001, \quad p(E') = 0.56006,$$

and hence, by Bayes's theorem,

$$p(T,E') = 0.62 \times .9/0.560.. = .996$$

$$p(H,E') = 0.9001 \times .6/0.560.. = .964.$$

Of course we expect the probabilities of T and H both to be substantially increased by the confirmation and we expect H to remain less probable than T. But it is interesting that the confirmation increases the probability of T much more than the refutation decreased it. One can conjecture that a theory could withstand a long succession of refutations of this sort (each depending on a different auxiliary hypothesis) punctuated by only an occasional confirmation, and its subjective probability still steadily increase on average. Detailed calculation indeed bears out this conjecture, which is again entirely in line with the views of Lakatos.

However Lakatos's position was little more than a bold conjecture in empirical sociology with arbitrary normative accretions. My Bayesian analysis brings out the underlying rationality of scientists who behave as Lakatos says they do. It also brings out the limitations of the Lakatosian methodology. For the Bayesian analysis yields nothing like these results if p(T) is initially assumed less than 1/2. There seems to be a negative moral here both for social scientists who might have hoped to draw comfort from the methodology of research programmes, and for those philosophers of science who deny that scientific theories can ever acquire respectable subjective probabilities.

I am grateful to Noretta Koertge and Michael Redhead for provoking me into analysing this problem in the first instance and for comments on earlier versions of the present paper.